

CLAIMS

We claim:

- 1 1. A method of compiling and accessing subject-specific information from a
2 computer network, the method comprising the steps of:
3 traversing links between sites on the computer network;
4 filtering the contents of each site visited to determine relevancy of content;
5 and
6 presenting information on each site deemed relevant for indexing.
- 1 2. The method according to Claim 1, further comprising the step of:
2 filtering the contents of a site at least a second time for relevancy, prior to the
3 step of presenting.
- 1 3. The method according to Claim 2, wherein at least one of said filtering steps
2 comprises the steps of:
3 presenting the contents to a human editor;
4 approving, by the human editor, if the contents are deemed relevant; and
5 disapproving, by the human editor, if the contents are not deemed relevant.
- 1 4. The method according to Claim 2, wherein at least one of said filtering steps
2 comprises the step of:
3 passing the contents of the site through a lexicon-based filter, the filter
4 comparing contents of the site with terminology found in the lexicon.
- 1 5. The method according to Claim 4, wherein the step of passing the contents of
2 the site through a lexicon-based filter comprises the steps of:
3 breaking up a web page corresponding to the site contents into component
4 parts; and

5 · comparing the contents of each component part with the lexicon.

1 6. The method according to Claim 5, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

3 assigning a weight to each component part based on a result of the step of
4 comparing; and
5 deeming the component part to be relevant if it achieves a high-enough
6 weight.

1 7. The method according to Claim 6, wherein the step of assigning a weight
2 comprises the steps of:

3 assigning a weight to each word, term, or expression in the component part
4 that matches a word, term, or expression in the lexicon, according to a weight associated with
5 the word, term, or expression; and
6 accumulating a sum of assigned weights, the sum forming the weight assigned
7 to the component part.

1 8. The method according to Claim 6, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

3 saving component parts deemed to be relevant and passing them to the
4 presenting step; and
5 discarding component parts deemed not to be relevant.

1 9. The method according to Claim 6, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

3 if at least one component part is deemed to be relevant, passing the web page
4 to the presenting step; and
5 if no component part is deemed to be relevant, discarding the web page.

10. The method according to Claim 4, wherein the step of passing the contents of the site through a lexicon-based filter comprises the step of:

comparing the contents of a web page corresponding to the site with the lexicon.

11. The method according to Claim 10, wherein the step of passing the contents of the site through a lexicon-based filter further comprises the steps of:

assigning a weight to the web page based on a result of the step of comparing; and

deeming the web page to be relevant if it achieves a high-enough weight.

12. The method according to Claim 11, wherein the step of assigning a weight comprises the steps of:

assigning a weight to each word, term, or expression in the web page that matches a word, term, or expression in the lexicon, according to a weight associated with the word, term, or expression; and

accumulating a sum of assigned weights, the sum forming the weight assigned to the web page.

13. The method according to Claim 11, wherein the step of deeming comprises the steps of:

saving the web page and passing it to the step of presenting if it achieves a high-enough weight; and

discarding the web page if it does not achieve a high-enough weight.

14. The method according to Claim 1, wherein the step of filtering the contents comprises the step of:

passing the contents of the site through a lexicon-based filter, the filter comparing contents of the site with terminology found in the lexicon.

1 15. The method according to Claim 14, wherein the step of passing the contents of
2 the site through a lexicon-based filter comprises the steps of:

3 breaking up a web page corresponding to the site contents into component
4 parts; and

5 comparing the contents of each component part with the lexicon.

1 16. The method according to Claim 15, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

3 assigning a weight to each component part based on a result of the step of
4 comparing; and

5 deeming the component part to be relevant if it achieves a high-enough
6 weight.

1 17. The method according to Claim 16, wherein the step of assigning a weight
2 comprises the steps of:

3 assigning a weight to each word, term, or expression in the component part
4 that matches a word, term, or expression in the lexicon, according to a weight associated with
5 the word, term, or expression; and

6 accumulating a sum of assigned weights, the sum forming the weight assigned
7 to the component part.

1 18. The method according to Claim 16, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

3 saving component parts deemed to be relevant and passing them to the
4 presenting step; and

5 discarding component parts deemed not to be relevant.

1 19. The method according to Claim 16, wherein the step of passing the contents of
2 the site through a lexicon-based filter further comprises the steps of:

if at least one component part is deemed to be relevant, passing the web page to the presenting step; and

if no component part is deemed to be relevant, discarding the web page.

20. The method according to Claim 14, wherein the step of passing the contents of the site through a lexicon-based filter comprises the step of:

comparing the contents of a web page corresponding to the site with the lexicon.

21. The method according to Claim 20, wherein the step of passing the contents of the site through a lexicon-based filter further comprises the steps of:

assigning a weight to the web page based on a result of the step of comparing;

and

deeming the web page to be relevant if it achieves a high-enough weight.

22. The method according to Claim 21, wherein the step of assigning a weight comprises the steps of:

assigning a weight to each word, term, or expression in the web page that matches a word, term, or expression in the lexicon, according to a weight associated with the word, term, or expression; and

accumulating a sum of assigned weights, the sum forming the weight assigned to the web page.

23. The method according to Claim 21, wherein the step of deeming comprises the steps of:

saving the web page and passing it to the step of presenting if it achieves a high-enough weight; and

discarding the web page if it does not achieve a high-enough weight.

24. The method according to Claim 14, further comprising the step of:

2 filtering the contents of a site at least a second time for relevancy, prior to the
3 step of presenting.

1 25. The method according to Claim 24, wherein the step of filtering the contents at
2 least a second time comprises the steps of:

3 presenting the contents to a human editor;
4 approving, by the human editor, if the contents are deemed relevant; and
5 disapproving, by the human editor, if the contents are not deemed relevant.

1 26. The method according to Claim 14, further comprising the step of:
2 replacing the lexicon with a lexicon corresponding to a different subject in
3 order to create a different subject-specific database.

1 27. The method according to Claim 1, further comprising the step of:
2 compiling a database of searchable relevant information.

1 28. The method according to Claim 27, further comprising the steps of:
2 permitting a user to enter a query; and
3 searching the database for information according to the query.

1 29. The method according to Claim 28, further comprising the step of:
2 displaying information found in said step of searching in a hierarchical format.

1 30. The method according to Claim 28, further comprising the step of:
2 determining a site ranking for each site associated with information found in
3 said searching step, where the determining is according to how interesting at least one of
4 authors and users of the computer network have found the site associated with the
5 information.

1 31. The method according to Claim 30, further comprising the step of:

2 displaying the results of the user query using the site ranking of each item of
3 [SAM1]information found in the search to determine an order in which the results are
4 displayed.

1 32. The method according to Claim 31, wherein the step of displaying the results
2 of the user query comprises the step of:

3 displaying the results of the user query in a hierarchical format according to
4 site ranking.

1 33. The method according to Claim 27, wherein the step of compiling a database
2 comprises the step of:

3 for each relevant site to be stored in the database, assigning a word score to
4 each word appearing on that site.[SAM3]

1 34. The method according to Claim 33, wherein the step of assigning word scores
2 comprises the steps of:

3 determining all sites found in the database that contain links to the site;
4 for each word on the site, assigning a word score for that word based at least
5 in part on its presence on each site containing a link to the site.

1 35. The method according to Claim 34, wherein the step of assigning a word score
2 for that word further comprises the step of increasing the word score for each site containing
3 a link to the site if the word appears in close proximity to the link.

1 36. The method according to Claim 33, wherein the step of assigning word scores
2 comprises the steps of:

3 determining all sites found in the database that contain links to the site; and
4 assigning a word score to each word on the site based at least in part on how
5 many sites linking to the site also contain the particular word.

1 37. The method according to Claim 36, wherein the step of assigning a word score
2 for that word further comprises the step of increasing the word score for each site containing
3 a link to the site according to the proximity of the word to the link.

1 38. The method according to Claim 33, further comprising the steps of:
2 entering a user query;
3 using the user query to search the database; and
4 computing a site ranking for each site associated with information found in
5 said searching step, the site ranking being computed based on said word scores.

1 39. The method according to Claim 38, wherein the step of computing a site
2 ranking comprises the steps of:
3 for each site associated with information found in said searching step,
4 summing the word scores for that site corresponding to words in the user query.

1 40. A computer-readable medium containing software implementing the method
2 as claimed in Claim 1.

1 41. A system for compiling and accessing information from a computer network,
2 the system comprising:
3 a processor; and
4 a computer-readable medium as claimed in Claim 40.

1 42. The method according to Claim 1, further comprising the step of:
2 monitoring a depth for each link, the depth being a reflection of relevance.

1 43. The method according to Claim 42, wherein the step of monitoring comprises
2 the steps of:
3 for a given site being visited, setting depths of any links leading from that site
4 to other sites to a depth of a link traversed to reach the given site;

5 if the given site is determined to be relevant in the filtering step, setting the
6 depths of the links leading from that site to zero; and

7 if the given site is determined not to be relevant in the filtering step,
8 incrementing the depths of the links leading from that site.

1 44. The method according to Claim 43, wherein the step of monitoring further
2 comprises the steps of:

3 comparing the incremented depths to a predetermined maximum depth value;
4 if the incremented depths exceed the predetermined maximum depth value,
5 discarding the links leading from the given site;
6 if the incremented depths do not exceed the predetermined maximum depth
7 value, traversing one of the links leading from the given site.

1 45. The method according to Claim 1, wherein said filtering step comprises the
2 steps of:

3 presenting the contents to a human editor;
4 approving, by the human editor, if the contents are deemed relevant; and
5 disapproving, by the human editor, if the contents are not deemed relevant.

1 46. A system that compiles and permits accessing of subject-specific information
2 from a computer network, the system comprising:

3 a host computer executing software from a computer-readable medium, the
4 software comprising:

5 a smart crawler for traversing the computer network;
6 a first filter, filtering out irrelevant sites, and permitting only relevant sites to pass;

7 and

8 an indexer indexing the relevant sites; and

memory, connected to the host computer, for storing indexed subject-specific information.

47. The system according to Claim 46, wherein said first filter comprises a lexicon-based filter.

48. The system according to Claim 47, wherein the system further comprises an interchangeable computer-readable medium on which is stored the lexicon for the lexicon-based filter, the lexicon containing subject-specific terminology.

49. The system according to Claim 46, wherein the software further comprises at least a second filter.

50. The system according to Claim 49, wherein the system further comprises a human-computer interface, and wherein at least one of said first filter and said at least a second filter comprises:

a presentation of relevant site information received from the smart crawler to a human editor via the human-computer interface; and
means for receiving input from the human editor, entered via the human-computer interface, as to whether or not to index and store the site in the memory.

51. The system according to Claim 49, wherein at least one of said first filter and said at least a second filter comprises a lexicon-based filter.

52. The system according to Claim 51, wherein the system further comprises an interchangeable computer-readable medium on which is stored the lexicon for the lexicon-based filter, the lexicon containing subject-specific terminology.

53. The system according to Claim 46, wherein the system further comprises a human-computer interface, and wherein said first filter comprises:

a presentation of relevant site information received from the smart crawler to a human editor via the human-computer interface; and

means for receiving input from the human editor, entered via the human-computer interface, as to whether or not to index and store the site in the memory.

54. A method of ranking the relevance of information stored in a database, the information comprising web pages, the method comprising the steps of:

computing and storing a word ranking for each word, except for stop words, found on each web page; and

in response to a user query, computing a site ranking for each web page found in response to the user query based on the word rankings.

55. The method according to Claim 54, wherein the step of computing a word ranking is performed according to how interesting at least one of authors and users of a computer network in which each web page is resident have found the web page.

56. The method according to Claim 54, wherein the step of computing a word ranking comprises the step of:

for each word, except stop words, on each web page, determining all web pages found in the database that contain links to the web page on which the word appears; and

assigning a word score for that word based at least in part on its presence on each web page containing a link to the web page on which that word appears, the word score constituting the word ranking for that word.

57. The method according to Claim 56, wherein the step of assigning a word score for that word further comprises the step of increasing the word score for each web page containing a link to the web page on which that word appears if the word appears in close proximity to the link.

58. The method according to Claim 54, wherein the step of computing a site ranking comprises the steps of:

3 for each web page found in response to the user query, summing the word
4 rankings for that web page corresponding to words in the user query.

1 59. A computer-readable medium containing software implementing the method
2 of Claim 54.